1-1-2012

Robust designs for poisson regression models

J M. Mcgree
Queensland University Of Technology

J A. Eccleston
University of Queensland

Robust designs for poisson regression models

Abstract

We consider the problem of how to construct robust designs for Poisson regression models. An analytical expression is derived for robust designs for first-order Poisson regression models where uncertainty exists in the prior parameter estimates. Given cer- tain constraints in the methodology, it may be necessary to extend the robust designs for implementation in practical experiments. With these extensions, our methodology con- structs designs which perform similarly, in terms of estimation, to current techniques, and offers the solution in a more timely manner. We further apply this analytic result to cases where uncertainty exists in the linear predictor. The application of this method- ology to practical design problems such as screening experiments is explored. Given the minimal prior knowledge that is usually available when conducting such experiments, it is recommended to derive designs robust across a variety of systems. However, in- corporating such uncertainty into the design process can be a computationally intense exercise. Hence, our analytic approach is explored as an alternative.

Keywords

robust, designs, poisson, regression, models

Publication Details

Mcgree, J. M. & Eccleston, J. A. (2012). Robust designs for poisson regression models. Technometrics, 54 (1), 64-72.

ROBUST DESIGNS FOR POISSON REGRESSION MODELS

J.M. McGree[1] and J.A. Eccleston[2]

[1]Queensland University of Technology
GPO Box 2434, Brisbane, QLD
Australia, 4001
james.mcgree@qut.edu.au

[2]University of Queensland
Hawken Dr, St. Lucia, QLD
Australia, 4072
j.eccelston@uq.edu.au

Abstract

We consider the problem of how to construct robust designs for Poisson regression models. An analytical expression is derived for robust designs for first-order Poisson regression models where uncertainty exists in the prior parameter estimates. Given certain constraints in the methodology, it may be necessary to extend the robust designs for implementation in practical experiments. With these extensions, our methodology constructs designs which perform similarly, in terms of estimation, to current techniques, and offers the solution in a more timely manner. We further apply this analytic result to cases where uncertainty exists in the linear predictor. The application of this methodology to practical design problems such as screening experiments is explored. Given the

[1]Author to whom correspondence should be addressed.

minimal prior knowledge that is usually available when conducting such experiments, it is recommended to derive designs robust across a variety of systems. However, incorporating such uncertainty into the design process can be a computationally intense exercise. Hence, our analytic approach is explored as an alternative.

Key words: Analytical solution; Canonical form; Compromise design; Average model; Poisson regression; Robust design.

1 Introduction

Optimal design for generalized linear models (GLMs) is complicated by the dependence of designs upon the model and parameter values, both of which are usually unknown *a priori*. This has been a major stumbling block for the development of optimal design for GLMs. However, recently there have been substantial advances in developing algorithmic approaches to generate robust designs which provide a means of addressing this complication. These advances have come from Woods, Lewis, Eccleston and Russell (2006), Dror and Steinberg (2006) and Gotwalt, Jones and Steinberg (2009). In order to apply these methods, uncertainty is described in the form of a probability distribution (or prior), and the objective is to then find a design that performs well over this uncertainty. These recent advances are applicable to GLMs in general with relatively complex linear predictor models with many covariates.

This paper is focussed on the robust design for Poisson models, and there has been some research into these specific models. Recent research by Russell, Woods, Lewis and Eccleston (2009), Rodriguez-Torreblanca and Rodriguez-Diaz (2007) and Wang, Myers, Smith and Ye (2006) are of particular interest. Wang, Myers, Smith and Ye (2006) investigate one and two variable cases with interactions and quadratic terms. They show how D-optimal designs depend on ratios of certain parameters, and derive upper bounds for these ratios. In addition, the performance of some 'standard' designs that appeal to practitioners was assessed. Their work (including previous papers) is concerned with applications to toxicology. Rodriguez-Torreblanca and Rodriguez-Diaz (2007) derived analytical solutions for D- and c-optimal designs for Poisson and negative binomial regression models. It was shown that for the Poisson case, D-optimal designs are invariant to the choice of intercept. Similarly, analytical results were derived for certain c-optimal designs. Russell, Woods, Lewis and Eccleston (2009) derived an

analytic solution for D-optimal designs for main effects Poisson models. This result generally eliminates the need to undertake computational and/or search algorithms to derive an optimal design. However, to produce designs robust with respect to the uncertainty in model parameters, an algorithmic approach was necessary.

Here we present an analytic result for the derivation of designs robust across a prior distribution on the parameter space meaning that computational and time expensive methods typically associated with robust designs are unnecessary. No such result has been presented in optimal design research for Poisson regression models, and as such this presents a substantial contribution to the design methodology. Our method may also prove useful for screening experiments where Poisson data arises, and we demonstrate the application of our analytical result through a practical example from Wu and Hamada (2000), see pages 563-573. The direct application of our methodology to practical design problems requires necessary conditions in order for our theoretical results to hold. We consider these conditions, and show how our methods can be easily extended for implementation in practice. Woods, Lewis, Eccleston and Russell (2006) introduced the idea of a compromise design, which is a robust design found by optimizing across many parameter sets simultaneously through the implementation of a compound criterion. In what follows, we define an 'average' model and show how it can be used to find an analytical solution for robust designs. The notion of compound or multiple objective criteria has been foreshadowed in other work. For further reading in this area, see Cook and Wong (1994), Clyde and Chaloner (1996), Huang and Wong (1998) and McGree, Eccleston and Duffull (2008).

The paper is outlined as follows. Initially methods are presented for deriving optimal and robust designs for Poisson models. A theorem is given which provides an analytical result for robust (saturated) designs for main effects models. It is shown by scaling up

the saturated design, our methods derive designs that have similar properties to those found via computational methods. An approach for the application of this methodology is proposed for cases where uncertainty also exists in the inclusion of covariates and therefore is directly applicable to screening experiments. Examples follow including the design of a seven-factor screening experiment from Wu and Hamada (2000) which demonstrates how to apply our methods in practice. Other examples are considered to compare computational approaches from the literature (Woods, Lewis, Eccleston and Russell (2006) and Gotwalt, Jones and Steinberg (2009)) with our simpler and more efficient techniques.

2 Background

Consider an experiment in which the ith observation on a response variable, y_i, has a Poisson distribution with rate λ_i dependent on p independent covariates $x_1, \ldots, x_p$ through the log-linear model:

$$\ln(\lambda_i) = \eta_i = f(\boldsymbol{x_i})^{\mathrm{T}}\boldsymbol{\beta} = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ji},\ i = 1, \ldots, n, \tag{1}$$

where $\boldsymbol{x_i} = (x_{1i}, \ldots, x_{pi})^{\mathrm{T}}$, $f(\boldsymbol{x_i}) = (1, \boldsymbol{x_i}^{\mathrm{T}})^{\mathrm{T}}$, $\beta_0, \ldots, \beta_p$ are unknown constants, and $\beta_j \neq 0$ for $j > 0$. For further details, see McCullagh and Nelder (1989).

Our aim is to design an experiment for the precise estimation of $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)^{\mathrm{T}}$. As is well-known, design for nonlinear models is based on initial or prior estimates of parameters of interest. Different estimates can lead to different designs and, as such, an optimal design for a given set of parameters is termed locally optimal. In practice, little may be known about the parameters (as this is generally the primary reason for

conducting the experiment). Initial or prior estimates are usually formed via expert opinion and/or based on any previous data collected. In any event, it would be wise to incorporate uncertainty about the parameters at the planning stage. Bayesian D-optimal methods have been suggested for dealing with the dependence of designs on initial parameter estimates. Unfortunately, this requires the evaluation of a generally intractable integral across some prior density. Chaloner and Larntz (1989), Woods, Lewis, Eccleston and Russell (2006) and Gotwalt, Jones and Steinberg (2009) have presented numerical methods for this evaluation/approximation. Another approach to form robust designs has come from Dror and Steinberg (2006) who apply clustering methods to local D-optimal designs. Russell, Woods, Lewis and Eccleston (2009) used this clustering idea, in conjunction with an analytical result, to yield robust designs.

2.1 Design

In general, a design is a collection/selection of points from a design space and is commonly expressed in one of two ways; as an approximate design or an exact design. Approximate designs are usually considered for theoretical development while actual experimentation requires an exact design. An approximate design $\xi \in \Xi$ in design space $\mathcal{X}$ with finite support is represented as

$$\xi = \left\{ \begin{array}{cccc} \boldsymbol{x_1} & \boldsymbol{x_2} & \dots & \boldsymbol{x_s} \\ \nu_1 & \nu_2 & \dots & \nu_s \end{array} \right\},$$

where $\boldsymbol{x_i} \in \mathcal{X}$, $\mathcal{X}$ is a compact subset of R^s, and the $\nu_i > 0$ are design weights that satisfy $\sum_{i=1}^{s} \nu_i = 1$ and represent the amount of experimental effort placed on the ith support point, $\boldsymbol{x_i}$. Exact designs are a special class of approximate designs. They have s

distinct sites, n runs and n_i runs at site i which defines an equivalent approximate design with weights n_i/n, for $i = 1, \ldots, s$. Russell, Woods, Lewis and Eccleston (2009) present a theorem for local approximate D-optimal designs for first-order Poisson regression models, and we summarize this result in the next section. The developed methodology in our paper is an extension of the work given by Russell, Woods, Lewis and Eccleston (2009) but focuses on exact designs.

2.2 Optimality

Under a Poisson regression model with linear predictor (1) and log link, the information matrix for ξ is

$$\begin{aligned} M(\xi, \boldsymbol{\beta}) &= \sum_{i=1}^{s} \nu_i w(\boldsymbol{x_i}) f(\boldsymbol{x_i}) f(\boldsymbol{x_i})^{\mathrm{T}} \\ &= \boldsymbol{X}^{\mathrm{T}} \boldsymbol{W} \boldsymbol{X}\,, \end{aligned}$$

where $w(\boldsymbol{x_i}) = \exp(\eta_i)$, $\boldsymbol{X} = (f(\boldsymbol{x_1}), \ldots, f(\boldsymbol{x_s}))^{\mathrm{T}}$ and $\boldsymbol{W} = \operatorname{diag}\{\nu_i w(\boldsymbol{x_i})\}_{i=1}^{s}$.

A locally D-optimal design, ξ^*, for a particular $\boldsymbol{\beta}$ is defined by

$$|M(\xi^*, \boldsymbol{\beta})|^{1/(p+1)} = \max_{\xi \in \Xi} |\boldsymbol{M}(\xi, \boldsymbol{\beta})|^{1/(p+1)}\,,$$

where the D-value of a given design ξ is $|\boldsymbol{M}(\xi, \boldsymbol{\beta})|^{1/(p+1)}$ and $p + 1$ is the number of parameters to be estimated.

The efficiency of a design ξ is then measured relative to ξ^* by

$$\{|\boldsymbol{M}(\xi,\boldsymbol{\beta})|/|\boldsymbol{M}(\xi^*,\boldsymbol{\beta})|\}^{1/(p+1)}. \tag{2}$$

In order to suppress the dependence of the design problem on $\boldsymbol{\beta}$, we follow Ford, Torsney and Wu (1992) and apply a linear transformation to $f(\boldsymbol{x_i})$ to obtain

$$f(\boldsymbol{z_i}) = \boldsymbol{B}f(\boldsymbol{x_i}),\ i = 1,\ldots,s, \tag{3}$$

where $\boldsymbol{z_i} = (z_{1i},\ldots,z_{pi})^{\mathrm{T}}$ belongs to the induced design space $\mathcal{Z}$,

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{B_{11}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B_{22}} \end{pmatrix},\ \boldsymbol{B_{11}} = \begin{pmatrix} 1 & 0 \\ \beta_0 & \beta_1 \end{pmatrix},$$

$\boldsymbol{B_{22}} = \mathrm{diag}\{\beta_2,\ldots,\beta_p\}$ and $\beta_j \neq 0$ $(j = 1,\ldots,p)$. It follows from equation (1) that $\eta_i = \left(\boldsymbol{B}^{-1}f(\boldsymbol{z_i})\right)^{\mathrm{T}}\boldsymbol{\beta} = \sum_{j=1}^{p} z_{ji}$. Let $\psi \in \Psi$ be a design measure over $\mathcal{Z}$. Then

$$\psi = \begin{Bmatrix} \boldsymbol{z_1} & \boldsymbol{z_2} & \ldots & \boldsymbol{z_s} \\ \nu_1 & \nu_2 & \ldots & \nu_s \end{Bmatrix}.$$

Let $\boldsymbol{e_j}$ denote the $p \times 1$ vector with ith element 1 if $i = j$, and 0 otherwise $(i, j = 1,\ldots,p)$.

Russell, Woods, Lewis and Eccleston (2009) derived the analytical result which states that a locally D-optimal design for the canonical first-order Poisson regression model with $\eta_i = (\boldsymbol{B}^{-1}f(\boldsymbol{z_i}))^T\boldsymbol{\beta} = \sum_{j=1}^{p} z_{ji}$, where $a_j \leq z_{ji} \leq b_j$, for a_j, b_j constants, and $b_j - a_j \geq 2$ $(j = 1,\ldots,p)$, is given by:

$$\psi^* = \left\{ \begin{array}{cccc} \boldsymbol{z_1}^* & \boldsymbol{z_2}^* & \dots & \boldsymbol{z_{p+1}}^* \\ 1/(p+1) & 1/(p+1) & \dots & 1/(p+1) \end{array} \right\}.$$

where $\boldsymbol{z_j}^* = \boldsymbol{b} - 2\boldsymbol{e_j}$, $j = 1, \dots, p$ and $\boldsymbol{z_{p+1}}^* = \boldsymbol{b}$, for $\boldsymbol{b} = (b_1, \dots, b_p)^T$.

This shows an analytical result to find a D-optimal design for first-order Poisson regression models for a fixed set of covariates and $\boldsymbol{\beta}$. It also shows that saturated designs, designs where the number of support points equals the number of parameters to be estimated, are D-optimal across the class of all designs.

There are implications when using this optimality result in practice. Firstly, the design is anchored at the point $\boldsymbol{z_{p+1}}^*$, in which all factors are set at the level that will give maximal expected response. To use this result, the experimenter needs to have a clear prior idea of which factors will increase and which will decrease the expected response. The second implication is that the design has a one-factor-at-a-time structure, with each factor moved 'in' from the anchor by an amount that reduces the expected response by a fixed amount. The design will not be able to estimate interactions or higher-order terms. Lastly, each factor must have a strong enough effect to achieve the required difference in the expected response. Later, this optimality result is extended to the derivation of robust designs, and again these implications require careful thought.

In order to find robust designs, computational approaches are generally needed. Russell, Woods, Lewis and Eccleston (2009) showed how this analytic result can be used with the clustering method of Dror and Steinberg (2006) to derive robust designs. Another approach has come from Woods, Lewis, Eccleston and Russell (2006), and we outline their approach next.

2.3 Compromise design

Following terminology given by Woods, Lewis, Eccleston and Russell (2006), a compromise D-optimal design over m alternative models is one which maximizes the following criterion:

$$\phi = |\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W_1}\boldsymbol{X}|^{\alpha_1/(p+1)} \times |\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W_2}\boldsymbol{X}|^{\alpha_2/(p+1)} \times \ldots \times |\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W_m}\boldsymbol{X}|^{\alpha_m/(p+1)}, \quad (4)$$

where α_k is the weight given to the kth alternative model and is such that $\sum_{k=1}^{m} \alpha_k = 1$.

Here the alternative models could represent a sample of prior estimates for $\boldsymbol{\beta}$, different link functions and/or different linear predictors. Woods, Lewis, Eccleston and Russell (2006) employed simulated annealing, a computationally intense and time consuming search algorithm (see Corana et al. (1987)), to maximize the above criterion. As an alternative to this and other computational methods, in the next section we present an analytic result for deriving such designs.

3 Robust design for Poisson regression models

The analytical result of Russell, Woods, Lewis and Eccleston (2009) is quite powerful. However, it only relates to a single model (a single vector $\boldsymbol{\beta}$). Here we derive an analytical result similar to that of Russell, Woods, Lewis and Eccleston (2009) but applicable to multiple sets of $\boldsymbol{\beta}$s from which we can obtain robust designs across multiple models very quickly. Although this new methodology is for $n = p + 1$, it is readily adaptable for deriving robust designs for $n > p + 1$. This is illustrated in the examples

that follow in Section 4. We also address the constraint of $b_j - a_j \geq 2$, as given by Russell, Woods, Lewis and Eccleston (2009).

3.1 Robust design with regard to initial parameter estimates

To set the scene, consider the case of m first-order alternative models, each with p covariates $\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_p}$. The models can be expressed as:

$$\begin{aligned} \ln(\boldsymbol{\lambda_1}) &= \boldsymbol{\eta_1} &&= \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta_1} \\ \ln(\boldsymbol{\lambda_2}) &= \boldsymbol{\eta_2} &&= \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta_2} \\ &\vdots \\ \ln(\boldsymbol{\lambda_m}) &= \boldsymbol{\eta_m} &&= \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta_m}. \end{aligned}$$

Suppose that the kth model has a pre-defined weighting α_k such that $\sum_{k=1}^{m} \alpha_k = 1$, corresponding to which initial estimates of $\boldsymbol{\beta}$ are believed to be more likely. We define an 'average' model, $\ln(\boldsymbol{\lambda_C})$, to be

$$\begin{aligned} \ln(\boldsymbol{\lambda_C}) &= \sum_{k=1}^{m} \alpha_k \boldsymbol{\eta_k} \\ &= \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\theta}, \end{aligned} \tag{5}$$

where $\boldsymbol{\theta}$ represents the prior parameter estimates for the average model.

The locally D-optimal design for the model in equation (5) has a design matrix $\boldsymbol{X}$ which maximizes

$$\phi_\theta = |\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W_\theta}\boldsymbol{X}|^{1/(p+1)}, \tag{6}$$

where

$$\boldsymbol{W_\theta} \;=\; \mathrm{diag}(\exp(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\theta})).$$

This leads to the following theorem. Given m alternative main effect models each having $p+1$ parameters, define a saturated design as one with $n = p+1$ experimental runs, then

Theorem 1 *The locally D-optimal saturated design for the average model is also the D-optimal saturated compromise design over the m models.*

Proof 1 *A saturated design means that $n = p+1 = q$, then $\boldsymbol{X}$ is a $q \times q$ matrix with rank q; the number of parameters in each model (including the average model). Given this and the fact that $\sum_{k=1}^{m} \alpha_k = 1$, the criterion in equation (4) can be re-expressed as:*

$$\begin{aligned}\phi &= |\boldsymbol{X}^{\mathrm{T}}|^{1/q} \times |\boldsymbol{W_1}|^{\alpha_1/q} \times |\boldsymbol{W_2}|^{\alpha_2/q} \times \ldots \times |\boldsymbol{W_m}|^{\alpha_m/q} \times |\boldsymbol{X}|^{1/q} \\ &= (|\boldsymbol{X}^{\mathrm{T}}| \times |\boldsymbol{W_c}| \times |\boldsymbol{X}|)^{1/q},\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{W_c} &= \operatorname{diag}\left\{(\exp(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta_1}))^{\alpha_1} \times \ldots \times (\exp(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta_m}))^{\alpha_m}\right\} \\
&= \operatorname{diag}\left\{(\exp(\boldsymbol{X}^{\mathrm{T}}\alpha_1\boldsymbol{\beta_1})) \times \ldots \times (\exp(\boldsymbol{X}^{\mathrm{T}}\alpha_m\boldsymbol{\beta_m}))\right\} \\
&= \boldsymbol{W_\theta}. \diamondsuit
\end{aligned}$$

Therefore, the consideration of the average model allows for the derivation for the compromise design. Given the average model is first order, results from Russell, Woods, Lewis and Eccleston (2009) can be applied directly to find the compromise design (analytically).

Numerical results from Dror and Steinberg (2006) and Woods, Lewis, Eccleston and Russell (2006) support the idea of optimizing an average (or centroid) model to form a robust design. Their work provided empirical evidence that the locally optimal design for the prior mean model of $\boldsymbol{\beta}$ (centroid or average model) is robust with respect to variations of $\boldsymbol{\beta}$ about the prior mean. Theorem 1 shows that this approach is optimal for saturated designs for first-order Poisson regression models.

3.2 Robust design with regard to initial parameter estimates and linear predictor

In this section, we conjecture that applying the average model approach will form efficient robust designs, not only when uncertainty exists in the initial parameter estimates, but also in the inclusion of covariates in the linear predictor. The reasoning behind this conjecture can be understood by first noting that the specific levels of each covariate for a (locally) D-optimal design are independent of the inclusion (and/or exclusion)

of other covariates in a main effects Poisson regression model. This can be seen by considering the locally D-optimal designs for the following two models.

$$\begin{aligned} \ln y_1 &= 2.5 + 2.5x_1 \\ \ln y_2 &= 2.5 + 2.5x_1 + 2.5x_2, \end{aligned}$$

where $x_{ji} \in [-1, 1]$, for $j = 1, 2$ and $i = 1, \ldots, n$.

The locally D-optimal designs for both models (respectively) are given below.

$$\xi_1^* = \begin{bmatrix} 0.2 \\ 1 \end{bmatrix}, \xi_2^* = \begin{bmatrix} 0.2 & 1 \\ 1 & 0.2 \\ 1 & 1 \end{bmatrix}.$$

Notice that the specific levels of x_1 are not affected by the presence of the other covariate in this main effects model. This is not specific to this example. Now, suppose we are interested in estimating parameters in the model which only has covariate x_1, based on each of the D-optimal designs. In order to put each design on an equal footing, ξ_1^* is replicated twice, and (column one of) ξ_2^* is replicated once. After this replication has occurred, the following two designs would (in each case) be used for the estimation of the model with only x_1 present

$$\xi_1 = \begin{bmatrix} 0.2 \\ 1 \\ 0.2 \\ 1 \\ 0.2 \\ 1 \end{bmatrix}, \xi_2 = \begin{bmatrix} 0.2 \\ 1 \\ 1 \\ 0.2 \\ 1 \\ 1 \end{bmatrix}.$$

We see that the design which was originally found for the model with only x_1 present is balanced, while the other design is not. With more covariates in the model, this lack of balance becomes more of an issue. However, given that the specific covariate levels remain unchanged, this led to the conjecture that the average model approach will yield efficient designs when uncertainty exists in the linear predictor.

The difficulty in proving this conjecture in the same manner as Theorem 1 was proved is that the compromise design will have a column for every covariate while the design for an alternative model may not, and hence the respective design matrix may not be square in parts of the criterion.

The following examples explore the estimation properties of designs found using Theorem 1 and the above conjecture.

4 Examples

In the examples that follow, the application of Theorem 1 and the conjecture to practical design problems is demonstrated. This application is not straightforward as we must

derive robust designs which have $n > p + 1$ experimental runs. To avoid computation, we propose replicating the saturated design as an approach to deal with this issue. However, one could consider computational approaches for deriving optimal choices for the remaining $n-(p+1)$ design points, conditional upon the $p+1$ support points given by the average model approach, see Waterhouse et al. (2009) for examples.

In the following, the first example considers the case where all models have the same covariates (Theorem 1). The second example extends this by allowing a different number of covariates in each alternative model so the conjecture can be explored. The final example compares our design approach with some of the computational methods previously referenced in this paper.

4.1 Example 1: Robustness in initial parameter estimates

Russell, Woods, Lewis and Eccleston (2009) considered a first order Poisson regression model and log link, where a robust design was required across the following parameter space, where $x_{ji} \in [-1, 1]$ for $i = 1, \ldots, n$ and $j = 1, \ldots, p$:

$$\beta_j = \begin{cases} [1, 1+\kappa] & \text{for } j = 1, 3, 5, 7, 9, \\ [-1-\kappa, -1] & \text{for } j = 2, 4, 6, 8, 10. \end{cases}$$

The robust design in Russell, Woods, Lewis and Eccleston (2009) was found by clustering local D-optimal design points based on the work of Dror and Steinberg (2006).

For this example, the saturated design contains 11 support points with equal weighting, but Russell, Woods, Lewis and Eccleston (2009) suggest the robust design should have 21 support points. Hence, Theorem 1 cannot be directly applied. To proceed, 10

support points from the saturated design need to be replicated. This will put our robust design on an equal footing with the one used in Russell, Woods, Lewis and Eccleston (2009). The choice of which 10 points to replicate can be made by considering the following theorem.

Theorem 2 *Given an exact design for a Poisson main effects model, taking an additional run of a support point will increase the determinant of the expected Fisher information by a factor of* $(1 + 1/r)$*, where* r *is the current number of design points at this site.*

Proof 2 *Let* $\boldsymbol{X}_n$ *denote a design matrix with* n *support points for a Poisson main effects model with* $n = p + 1$ *parameters. Suppose the additional run is taken at the* i*th support point* $f(\boldsymbol{x}_i)$*, denote the design matrix with* $n + 1$ *runs as* $\boldsymbol{X}_{n+1}$*. Adopt similar notation for the weight matrix* $\boldsymbol{W}$ *where* $w_{ii} = r_i \exp(f(\boldsymbol{x}_i)^{\mathrm{T}}\boldsymbol{\beta})$*, for* $i = 1, \ldots, n$*, where* r_i *is the number of runs at the* i*th support point. Consider the following.*

$$\begin{aligned}
\boldsymbol{X}_{n+1}^{\mathrm{T}}\boldsymbol{W}_{n+1}\boldsymbol{X}_{n+1} &= (\boldsymbol{X}_n^{\mathrm{T}}, f(\boldsymbol{x}_i)) \begin{bmatrix} \boldsymbol{W}_n & 0 \\ 0 & \exp(\eta_i) \end{bmatrix} \begin{pmatrix} \boldsymbol{X}_n \\ f(\boldsymbol{x}_i)^{\mathrm{T}} \end{pmatrix} \\
&= \boldsymbol{X}_n^{\mathrm{T}}\boldsymbol{W}_n\boldsymbol{X}_n + f(\boldsymbol{x}_i)\exp(\eta_i)f(\boldsymbol{x}_i)^{\mathrm{T}} \\
|\boldsymbol{X}_{n+1}^{\mathrm{T}}\boldsymbol{W}_{n+1}\boldsymbol{X}_{n+1}| &= |\boldsymbol{X}_n^{\mathrm{T}}\boldsymbol{W}_n\boldsymbol{X}_n + f(\boldsymbol{x}_i)\exp(\eta_i)f(\boldsymbol{x}_i)^{\mathrm{T}}|.
\end{aligned}$$

Consider the matrix

$$\begin{bmatrix} \boldsymbol{X}_n^{\mathrm{T}}\boldsymbol{W}_n\boldsymbol{X}_n & -f(\boldsymbol{x}_i) \\ \exp(\eta_i)f(\boldsymbol{x}_i)^{\mathrm{T}} & 1 \end{bmatrix},$$

and the following result for partitioned matrices

$$\begin{aligned} \det \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} &= |\boldsymbol{A}||\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}| \\ &= |\boldsymbol{D}||\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}|, \end{aligned}$$

provided $\boldsymbol{A}$ *and* $\boldsymbol{D}$ *are nonsingular matrices.*

From above, we have

$$|\boldsymbol{X}_n^{\mathrm{T}}\boldsymbol{W}_n\boldsymbol{X}_n + f(\boldsymbol{x}_i)\exp(\eta_i)f(\boldsymbol{x}_i)^{\mathrm{T}}| = |\boldsymbol{X}_n^{\mathrm{T}}\boldsymbol{W}_n\boldsymbol{X}_n||1 + f(\boldsymbol{x}_i)^{\mathrm{T}}\exp(\eta_i)(\boldsymbol{X}_n'\boldsymbol{W}_n\boldsymbol{X}_n)^{-1}f(\boldsymbol{x}_i)|. \tag{7}$$

Now $(\boldsymbol{X}_n^{\mathrm{T}}\boldsymbol{W}_n\boldsymbol{X}_n)^{-1} = \boldsymbol{X}_n^{-1}\boldsymbol{W}_n^{-1}(\boldsymbol{X}_n^{\mathrm{T}})^{-1}$*, and given that* $f(\boldsymbol{x}_i)$ *is a row of* $\boldsymbol{X}_n$*, it follows that*

$$f(\boldsymbol{x}_i)^{\mathrm{T}}\exp(\eta_i)(\boldsymbol{X}_n^{\mathrm{T}}\boldsymbol{W}_n\boldsymbol{X}_n)^{-1}f(\boldsymbol{x}_i) = 1/r_i.$$

Therefore,

$$|\boldsymbol{X}_{n+1}^{\mathrm{T}}\boldsymbol{W}_{n+1}\boldsymbol{X}_{n+1}| = (1 + 1/r_i)|\boldsymbol{X}_{n}^{\mathrm{T}}\boldsymbol{W}_{n}\boldsymbol{X}_{n}|. \diamondsuit$$

Note that a similar expression to equation (7) has been used as an updating formula for design search algorithms, see Atkinson, Donev and Tobias (2007) (pg: 174) and Woods (2010).

The above result says to maximize the determinant of the expected Fisher information matrix, the replication should be kept as even as possible amongst all support points. That is, if all support points of a given design have been replicated the same number of times, then the choice for replicating the next support point is arbitrary. However, if the number of replicates of each support point for a given design are not equal, then one should replicate a support point with the least runs.

Now returning to the example, 1000 simulations of the vector $\boldsymbol{\beta}$ were generated by selecting each parameter value from the relevant uniform distribution, and the robust design was found by giving each $\boldsymbol{\beta}$ a weight of $\alpha_k = 1/1000$, for $k = 1, \ldots, m$. The efficiencies of this design relative to the 1000 D-optimal designs (where $n = 21$) for the various values of $\boldsymbol{\beta}$ were calculated. Table 1 provides a summary of the results.

Table 1: Comparison between the analytical solution and the clustering algorithm used in Russell, Woods, Lewis and Eccleston (2009).

Method	Statistic	$\kappa = 1$	$\kappa = 2$	$\kappa = 3$
Russell, Woods, Lewis and Eccleston (2009)	Median D-eff.	0.936	0.877	0.748
	Minimum D-eff.	0.895	0.803	0.633
Average model approach (Theorem 1)	Median D-eff.	0.966	0.923	0.887
	Minimum D-eff.	0.928	0.840	0.756

From Table 1, the analytical solution provides larger median and minimum D-efficiencies than the clustering method for each value of κ, and is obviously quicker. This shows the benefits for considering our methodology when constructing robust designs.

For the case when $n > p+1$, the choice of n is unclear. In this example, Russell, Woods, Lewis and Eccleston (2009) used the Bayesian Information Criterion (see Fraley and Raftery (2002)) to determine $n = 21$. Dror and Steinberg (2006) constructed examples which show how the minimum and median D-efficiencies of robust designs may be significantly increased by the inclusion of more (than $p+1$) design points. Indeed, they suggest a method for determining an appropriate choice of n. Their approach can be used in conjunction with Theorem 1 to quickly construct robust designs.

4.2 Example 2: Uncertainty in initial parameter estimates and linear predictor

Wu and Hamada (2000) consider a screening experiment which yields count data. There are seven covariates ($x_{ji} \in [-1, 1]$) in consideration and it is unknown whether some or all of them are influential in explaining the response. We assume that expert opinion can advise whether the covariates act positively or negatively on the response (if influential) but also allow for the case where the initial parameter estimate could be zero. Therefore, there exists uncertainty in the linear predictor and the initial parameter estimates. The uncertainty in the linear predictor is significant. Given that some or all covariates could be influential, there exists the full main effects model, 7 one and six factor models, 21 two and five factor models and 35 three and four factor models. That is, a total of 127 (equally weighted) alternative models with uncertainty in parameter estimates needs to be considered. The prior parameter ranges (uniformly distributed) for all alternative

models are given below where the sign was taken from the analysis given in Wu and Hamada (2000).

$$\boldsymbol{\beta} \in \begin{bmatrix} 0 & 0 & -5 & 0 & -5 & -5 & -5 & -5 \\ 5 & 5 & 0 & 5 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

For this example, we evaluate the performance of three designs; a compromise design found using the methods of Woods, Lewis, Eccleston and Russell (2006), our average model approach and a 2^{7-3}_{IV} fractional factorial design (which has 16 support points) given in Wu and Hamada (2000).

To search for a compromise design, a random sample of twenty $\boldsymbol{\beta}$ values was drawn from the above parameter ranges. Simulated annealing was used to find the robust design (with $n = 16$ runs) which maximized the sum of the logarithm of D-values given under each alternative model with each random $\boldsymbol{\beta}$. Each of the 127 alternative models will have different linear predictors representing the uncertainty around which covariate/s should appear in the model. The sum of the logarithm of D-values was used instead of the product of D-values to avoid numerical problems (very large criterion values). Given there are 127 alternative (structural) models and uncertainty in the $\boldsymbol{\beta}$s represented by a sample of twenty, calculating the criterion value for a given design requires adding up 2540 log determinants. Therefore, maximizing this criterion with respect to the choice of design using, for example, simulated annealing, is therefore computationally intense.

To find a robust design using our average model approach, the weighted mean of each parameter was calculated over the prior parameter space and each alternative model (centroid of the parameter space for the full model). Given there exists uncertainty in the linear predictor, not all parameters may appear in each rival model. When this

occurs, the prior estimate for the parameter is calculated by ignoring the corresponding prior model weights α, and re-scaling the remaining αs such that they sum to one. This forms the average model, and, in this example, the model will have 8 parameters. The D-optimal saturated design for this model will therefore have 8 support points, and hence all are replicated to yield a 16 run design. The computation involved here is minimal.

The performance of the three above designs were evaluated by randomly and uniformly drawing 1000 parameter vectors from the above intervals and calculating D-efficiencies of all three designs with respect to the actual D-optimal design under all 127 alternative models. Figure 1 shows the median and minimum D-efficiencies for this evaluation (over 1000 parameter sets).

The superiority of the compromise design and the average model approach over the fractional factorial is obvious. This is highlighted by the median and minimum D-efficiencies generally being greater than those given by the fractional factorial. We also see that the compromise design and average model approach have different estimation properties. The D-efficiencies of the compromise design seem to decrease as the number of factors in the models increases. This is the opposite for the design found using an average model approach with D-efficiencies increasing with more factors. This difference may be attributed to the poor balance of designs for models with a small number of factors (as discussed in Section 3.2). If we consider the compromise design as the benchmark, the average model approach performs well across all models and provides support to the conjecture that our analytic result can be used to find efficient robust designs for screening type experiments.

We highlight that this example shows how to deal with uncertainty in initial parameter estimates containing a value of zero. The condition that $b_j - a_j \geq 2$ from Russell, Woods,

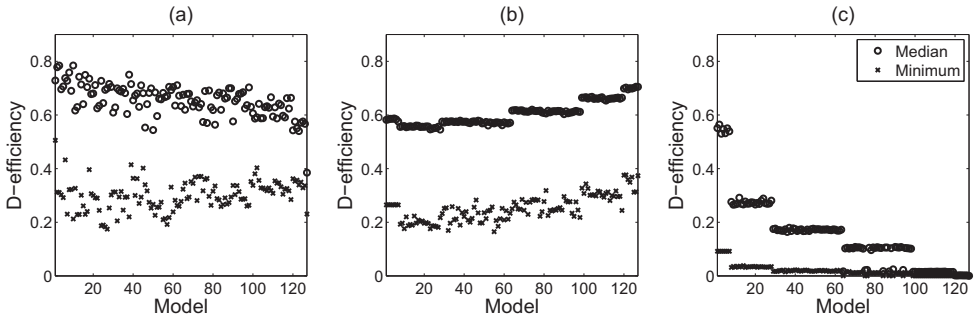

Figure 1: Median and minimum D-efficiencies for each of the 127 alternative models (ordered from one factor models to the seven factor main effects model) for the (a) compromise D-optimal design (b) D-optimal design for the average model and (c) orthogonal design from Wu and Hamada (2000).

Lewis and Eccleston (2009) is placed on the average model, not the alternative/rival models. Hence, uncertainty in this form can be accounted for.

4.3 Example 3: Simulation study

Both of the previous examples have been selected and constructed such that the condition $b_j - a_j \geq 2$ from Theorem 1 holds. The question remains, what do to when this condition does not hold for the average model? For such cases, we advise still forming the average model, and then searching for the (locally) optimal design for this specific model. We note that this requires an optimization algorithm, but will be as fast as a search for a single local optimal design providing significant savings in regard to computing times. Through this example then, we aim to provide empirical evidence to support the 1.) implementation of the average model approach and 2.) use of replication as efficient methods of designing experiments described by Poisson regression models. Let us suppose there are four independent variables each with $x_{ji} \in [-1, 1]$ $(j = 1, \ldots, p, i = 1, \ldots, n)$ that are thought to affect the response/distribution of a Poisson random variable. Consider prior knowledge about the models to be defined by

the different parameter ranges found in Table 2.

Three approaches for the derivation of robust optimal designs are compared. The first approach is described in Woods, Lewis, Eccleston and Russell (2006). Again, a random sample of parameters is drawn, and simulated annealing was employed to find a compromise design. The second approach was given by Gotwalt, Jones and Steinberg (2009) who gave an algorithm to find robust designs quickly. The algorithm can be found in JMP, version 8 and this package was used to find the robust designs here. The last approach, as described in this paper, forms the average model, and simply constructs designs given these point estimates of parameters.

The parameter ranges $B_1 - B_3$ were considered in Woods, Lewis, Eccleston and Russell (2006), and are such that Theorem 1 is not directly applicable. Therefore, for these parameter spaces, the D-optimal criterion for the average model was optimized using simulated annealing. Ranges $B_4 - B_6$ were arbitrarily defined, and notably allow for the direct application of Theorem 1. As such, the optimal design for the average model was derived analytically, and replicated appropriately. Hence, we aim to compare (fully) computational approaches with the methods proposed in this paper.

Table 2: Ranges for each model parameter for the parameter spaces B_k, for $k = 1, \ldots, 6$.

	Parameter space					
Parameter	B_1	B_2	B_3	B_4	B_5	B_6
β_0	$[-3, 3]$	$[-1, 1]$	$[-3, 3]$	$[0, 3]$	$[-1, 4]$	$[-3, 0]$
β_1	$[-2, 4]$	$[0, 2]$	$[4, 10]$	$[-2, 4]$	$[0, 2]$	$[4, 10]$
β_2	$[-3, 3]$	$[-1, 1]$	$[5, 11]$	$[1, 3]$	$[-1, 3]$	$[5, 11]$
β_3	$[0, 6]$	$[2, 4]$	$[-6, 0]$	$[1, 6]$	$[2, 4]$	$[-6, 0]$
β_4	$[-2.5, 3.5]$	$[-0.5, 1.5]$	$[-2.5, 3.5]$	$[-2.5, -3.5]$	$[-5.5, 1.5]$	$[0, 3.5]$

Before designs can be derived, a decision about the number of experimental runs needs to be made. Approaches from the literature are available (Bayesian Information Cri-

terion or a clustering approach), but we simply chose two arbitrarily, one small value ($n = 6$) and one large value ($n = 24$). After each design was found, a comparison study was performed. This involved randomly generating 20,000 models from each parameter space, and evaluating the D-values for each of the three robust designs for both choices of n. Relative D-efficiencies were calculated and shown in the form of (empirical) cumulative density plots, see Figures 2 and 3 for $n = 6$ and 24, respectively.

Relative D-efficiencies were calculated for robust designs found by the methods of Woods et al. (2006) and Gotwalt et al. (2009) compared with designs derived from our average model approach. Therefore, relative D-efficiencies less than one indicate the average model approach is performing 'better' than the two computational approaches (and the reverse is true for relative D-efficiencies greater than one). Upon inspecting Figure 2, we see that all plots (except one) cross the line of unity at 0.5 or higher, indicating that for the majority of simulations, the average model approach produced a more efficient design. The only exception can be seen in Figure 2e where the plot crosses just below 0.5 indicating that the approach of Gotwalt et al. (2009) produced a more efficient design in this case. In Figure 2b, only the plot relating to the approach of Gotwalt et al. (2009) is visible. This is because the design derived via methods from Woods et al. (2006) has a near identical performance to the robust design given by the average model approach. With reference to Figure 3, the average model approach again performs well. This is particularly noticeable for Figure 3f where both plots cross the line of unity at 0.6 or higher. There are parameter spaces where the average model approach does not perform as well as the two computational methods. This shows that our approach is not optimal (for $n > p + 1$ and $b_j - a_j < 2$), but shows how efficient robust designs can be formed with minimal computation.

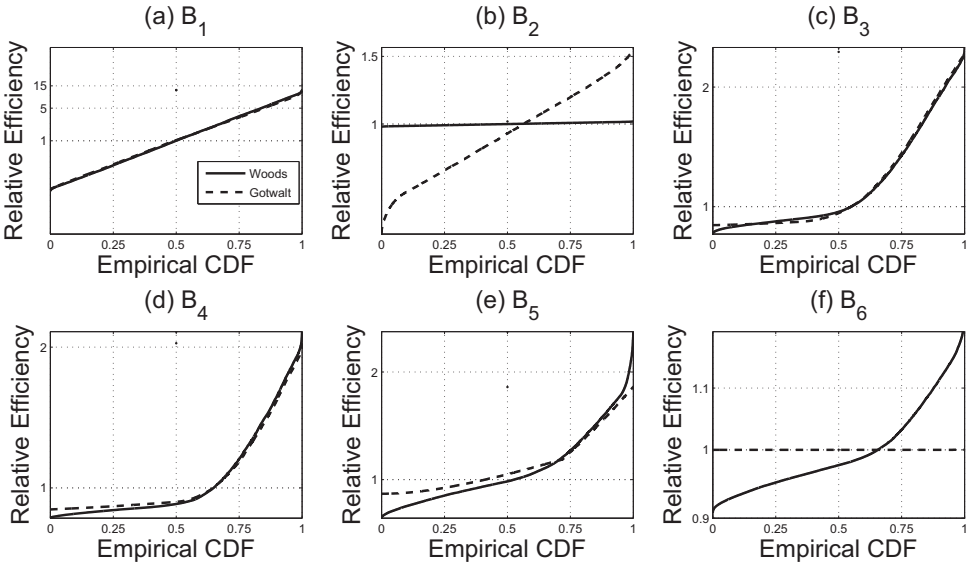

Figure 2: Empirical cumulative density plots for the relative D-efficiencies of designs found via the approach of Woods et al. (2006) and Gotwalt et al. (2009) compared with the average model approach for parameter spaces $B_1 - B_6$ where $n = 6$.

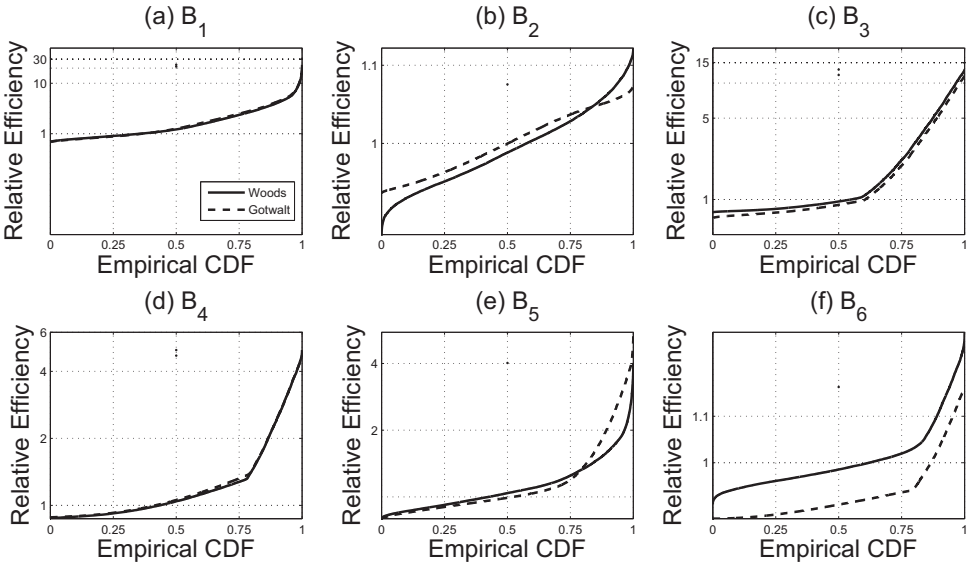

Figure 3: Empirical cumulative density plots for the relative D-efficiencies of designs found via the approach of Woods et al. (2006) and Gotwalt et al. (2009) compared with the average model approach for parameter spaces $B_1 - B_6$ where $n = 24$.

5 Conclusion

The results presented in this paper provide methods to derive robust designs for Poisson regression models with respect to uncertainty in the parameter vector $\boldsymbol{\beta}$ and the inclusion of covariates. They are applicable to many experimental situations and, generally being analytical, they naturally require minimal computation. Although Theorem 1 imposes the restriction that $n = p + 1$, as demonstrated, cases where $n > p + 1$ can be easily dealt with via replication. In general, little may be known about the regression coefficients and the presence or absence of a covariate in a model, and therefore it is important to consider such uncertainty in the design process. An example of this is in screening experiments, where our methods would be very applicable and useful.

Our methodology suggests that replication is beneficial in forming robust designs for Poisson regression models. The results from Examples 1 and 3 support this. Despite robust designs found via optimization algorithms yielding more than $p + 1$ support points, evaluations of these designs revealed either higher D-efficiencies (Example 1) or similar D-efficiencies (Example 3) for designs constructed via replication. We acknowledge that this may be somewhat counter-intuitive for the general construction of robust designs. One would feel that extra support points should provide greater robustness against, for example, uncertainty in parameter estimates. We do not argue against this point in general, but believe that this is not strictly so for robust designs for first-order Poisson regression models.

Robust designs derived from the average model approach for first order Poisson regression models will have no ability to estimate quadratic or higher-order terms, and in fact will only be able to estimate particular interactions via replication. If such terms are believed to be important, then these should be considered when designing the experi-

ment. We note that the proof of Theorem 1 does not assume any specific form of the linear predictor. No restriction was used to enforce a main effects model. Hence, it is of interest to investigate the performance of our methodology for models that include interactions, quadratic terms, etc. There are many issues to address. For example, unlike in the main effects case, there is no guarantee that the saturated D-optimal design is actually D-optimal across the class of all designs. Moreover, the specific levels of each covariate are likely to change with the inclusion or exclusion of high-order terms leading to further issues surrounding model uncertainty. Computational approaches are available for models that include higher-order terms, but may prove too time costly to run in practice, particularly with a large number of covariates and an uncertain linear predictor. The average model approach may prove useful.

Our approach allows robust designs to be quickly derived for complex situations with uncertainty in the parameter space and linear predictor. The alternatives are computational approaches which may take considerable time. The benefits of such computational methods were explored here and showed minimal reward. In conclusion, our approach should be used with caution and, as always, the performance of experimental designs should be investigated prior to data collection.

Acknowledgment We acknowledge discussions with Dave Woods, University of Southampton, U.K. and Ken Russell, Charles Sturt University. The authors also wish to thank the Editor, Associate Editor and both referees for their helpful comments throughout the review process.

References

Atkinson, A., Donev, A. D. and Tobias, R. D. (2007) *Optimum Experimental Designs, with SAS*. Oxford University Press 2nd edn.

Chaloner, K. and Larntz, K. (1989) Optimal Bayesian design applied to logistic regression experiments. *Journal of Statistical Planning and Inference* **21**, 191–208.

Clyde, M. and Chaloner, K. (1996) The equivalence of constrained and weighted designs in multiple objective design problems. *Journal of the American Statistical Association* **91**, 1236–1244.

Cook, R. D. and Wong, W. K. (1994) On the equivalence of constrained and compound optimal designs. *Journal of the American Statistical Association* **89**, 687–692.

Corana, A., Marchesi, M., Martini, C. and Ridella, S. (1987) Minimizing multimodal functions of continuous variables with the 'simulated annealing' algorithm. *ACM Transactions on Mathematical Software* **13**, 262–280.

Dror, H. A. and Steinberg, D. M. (2006) Robust experimental design for multivariate generalized linear models. *Technometrics* **48**, 520–529.

Ford, I., Torsney, B. and Wu, C. F. J. (1992) The use of a canonical form in the construction of locally optimal designs for nonlinear problems. *Journal of the Royal Statistical Society, Series B (Methodological)* **54**, 569–583.

Fraley, C. F. and Raftery, A. E. (2002) Model-based clustering, discriminant analysis and density estimation. *Journal of the American Statistical Association* **97**, 611–631.

Gotwalt, C. M., Jones, B. A. and Steinberg, D. M. (2009) Fast computation of designs robust to parameter uncertainty for nonlinear settings. *Technometrics* **51**, 88–95.

Huang, Y. C. and Wong, W. K. (1998) Multiple-objective optimal designs. *Journal of Biopharmaceutical Statistics* **8**, 635–643.

McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*. Chapman and Hall 2nd edn.

McGree, J. M., Eccleston, J. A. and Duffull, S. B. (2008) Compound optimal design criteria for nonlinear models. *Journal of Biopharmaceutical Statistics* **18**, 646–661.

Rodriguez-Torreblanca, C. and Rodriguez-Diaz, J. M. (2007) Locally D- and c-optimal designs for Poisson and negative binomial regression models. *Metrika* **66**, 161–172.

Russell, K. G., Woods, D. C., Lewis, S. M. and Eccleston, J. A. (2009) D-optimal designs for Poisson regression models. *Statistica Sinica* **19**, 721–730.

Wang, Y., Myers, R. H., Smith, E. P. and Ye, K. (2006) *d*-optimal designs for Poisson regression models. *Journal of Statistical Planning and Inference* **136**, 2831–2845.

Waterhouse, T. H., Eccleston, J. A. and Duffull, S. B. (2009) Optimal design criteria for discrimination and estimation in nonlinear models. *Journal of Biopharmaceutical Statistics* **19**, 386–402.

Woods, D. C. (2010) Robust designs for binary data: applications of simulated annealing. *Journal of Statistical Computation and Simulation* **80**, 29–41.

Woods, D. C., Lewis, S. M., Eccleston, J. A. and Russell, K. G. (2006) Designs for generalized linear models with several variables and model uncertainty. *Technometrics* **48**, 284–292.

Wu, C. F. J. and Hamada, M. (2000) *Experiments: Planning, analysis and parameter design optimization*. John Wiley & Sons, Inc. 1st edn.